

**Tomoko Imai\***

Department of Advanced  
Interdisciplinary Studies  
Graduate School of Engineering  
University of Tokyo  
Tokyo, Japan

**Dairoku Sekiguchi****Masahiko Inami†****Naoki Kawakami****Susumu Tachi**

Department of Information Physics  
and Computing  
Graduate School of Information  
Science and Technology  
University of Tokyo  
Tokyo, Japan

\*Current affiliation:

UDIT, Inc.  
Kanagawa, Japan  
tomoko@audit.jp

†Current affiliation:

Department of Mechanical  
Engineering and Intelligent Systems  
The University of Electro-  
Communications  
Tokyo, Japan

# Measuring Gaze Direction Perception Capability of Humans to Design Human Centered Communication Systems

---

**Abstract**

We describe experiments designed to measure gaze direction perception capability of humans under face-to-face and display mediated conditions. Gaze perception capability was determined by means of the absolute values of the pitch differences between a looker's actual regards and participants' judgments. We compared the capability under face-to-face, stereoscopic image, and monoscopic image conditions. On average, participants perceived the looker's gaze direction most accurately under the face-to-face condition. As expected, the accuracy under the stereoscopic image condition was higher than the results obtained under the monoscopic image condition. However, individual data did not follow the expected order and our exploratory experiments showed that participants with narrower interpupillary distance than the distance between two stereo cameras had difficulty in judging gaze directions. We also found that the perception of the pitch component of gaze direction is affected by gaze transmission methods but the yaw component is robust and is not affected by the transmission conditions.

**I Introduction**

If a communication system can transmit and receive all the information exchanged between people, the system can recreate natural face-to-face communication. So far, no system can transmit all the information exchanged between people accurately. Since each system has strengths and weaknesses, designers and users should be able to weigh these with respect to cost and the reliability of information that the system can convey. If the quality of the exchanged information can be measured quantitatively, consumers can compare communication systems objectively and researchers can evaluate their newly developed communication systems quantitatively to find out necessary improvements. In this paper, we compare gaze transmission capability of display systems by measuring accuracy of human perception of gaze directions under stereoscopic display mediated, monoscopic display mediated, and face-to-face conditions. The measurement results can be used to quantify similarity between the technology mediated remote communication and the face-to-face communication because the gaze has been regarded as an important component of face-to-face communication. The results are also useful to select appropriate displays to design new communication systems because people can select

displays based on price, usability, and the similarity indexes that show how far displays can recreate face-to-face communication.

Gaze has been regarded in many research fields as an important component of non verbal cues and studied in many fields. It is reported that humans acquire the capability of the joint visual attention before the age of 14 months and use this information to interact with others throughout their lives (Scaife & Bruner, 1975). The unique horizontally elongated morphology of the human eyes is believed to be evolved to communicate gaze signals efficiently (Kobayashi & Kohshima, 1997). Also, humans perceive gazes as shifts relative to an eye size rather than absolute (retinal) spatial parameters (Sato & Matsuzaki, 2000). In the field of psychophysics, gaze has been investigated as an important social cue. Gibson measured the relation between the gaze direction toward a participant and the sensation of being looked at. In the experiment, the looker and a participant sat 200 cm apart and a looker fixated on one of the horizontally aligned targets placed behind the participants, who would signal verbally when they felt they were being looked at (Gibson & Pick, 1963). He showed that participants felt not being looked at when the looker's gaze direction was deviated from a nose more than 9 cm, which is  $2.8^\circ$  from the looker's position. He also reported that judged targets were shifted into opposite direction of the head rotation when the looker rotated her head and looked askance at targets. Cline (1967) and Antsis, Mayhen, and Morley (1969) extended Gibson's experiment by measuring the relation between the looker's gaze directions and the participants' judged directions. In the experiment, participants judged not only whether they were looked at, but also where the looker was looking. Cline reported the same shift of the judged location in the opposite direction to the looker's head orientation but the shift was in the same direction as the head orientation when the head and target orientations were parallel. He also showed that the standard deviation is 1.55 cm at a distance of 122 cm and this amount is smaller than the Gibson's result. These research results suggest a complicated influence of head orientation on gaze direction perceptions. Anstis et al. showed that the yaw angles were overestimated by reporting a regression line for the perceived gaze yaw an-

gles as  $y = 1.50x - 0.05$ , where  $x$  is the yaw component of the looker's gaze direction and  $y$  is the yaw component of participants' judgment. They also conducted the same experiment using a TV screen and showed that participants can judge the looker's gaze direction with an accuracy of 4 cm when their distance is 200 cm. The result suggested that information on at whom the looker is looking can be transmitted using a monoscopic image on a TV screen because Gibson showed that people feel being looked at when the looker's gaze direction was deviated from a nose less than 9 cm. Since the purpose of these experiments was to investigate gaze toward face areas, the targets were arranged around the height of the participants' faces. However, the gaze directed toward working areas is also important because people look very little at faces when there are relevant objects or an interesting background to look at (Argyle & Cook, 1976).

In the field of computer science, gaze has been studied as one of the nonverbal cues that should be transmitted by communication systems. The influence of gaze on communication has been measured through human behavior observations (Fish, Kraut, & Chalfonte, 1990; Ishii, Kobayashi, & Jonathan, 1993), task performance measurements (Taylor & Rowe, 2000) and questionnaires (Vertegaal, 1999; Garau, Slater, Bee, & Sasse, 2001; Lee, Badler, & Badler, 2002). There are also efforts to use psychophysical experiments to assess communication systems. Anstis et al. (1969) measured the yaw component of perceived gaze directions and Tachi and Arai (1997) measured users' horopter curves to design and tune binocular displays to individual users (Tachi & Arai, 1997).

There have been no investigations to date (to the authors' knowledge) on how accurately people recognize collaborators' gaze onto desk spaces, and that compared the accuracy of gaze direction perception between the face-to-face and display mediated conditions. Since judging gaze toward a working area is necessary to know the partner's focus of attention, it is helpful for collaborators if communication systems can transmit gaze information toward working areas naturally. To design appropriate display technologies to support distant collaborations, it is necessary to know the accuracy of the gaze perception of humans under face-to-face

conditions and how the information is modified by different display technologies. We have extended the gaze direction perception measurement method developed by Anstis et al. (1969) to measure human perception of gaze directions toward desk spaces to answer the following questions: How is the visual information exchanged in face-to-face communication modified when a communication system is introduced? Is it possible to measure the difference quantitatively? Is it possible to use the result of the measurements to design and compare new communication systems?

### 1.1 Outline

Transmitting nonverbal cues is regarded as crucial to enable face-to-face like communication between remote sites (Fish, Kraut, & Root, 1992; Lee et al., 2002). A partner's gestures, head movements, eye contact, and gaze directions are used to shift the focus of attention coherently (Ishii et al., 1993). Our study focuses on one of the cues, gaze, and especially gaze toward desk spaces because we often try to understand collaborators' focus of attentions by judging gaze toward working areas but this kind of gaze has not been studied well compared to gazes toward a face area. The goal of this study is to measure the accuracy of human perception of gaze directions under the face-to-face condition to identify the natural gaze communication state and to compare this result to gazes mediated by monoscopic and stereoscopic displays. In this paper, a person who looks at targets is called "a looker" following naming conventions used in previous papers (Gibson & Pick, 1963; Cline, 1967; Anstis et al., 1969).

In the first experiment, we investigated how accurately people understand a partner's gaze directions toward a working area. The purpose of the experiment is to identify indexes that can represent a natural communication state so that display mediated communication can be evaluated quantitatively.

The second experiment measured gaze direction perception using a monoscopic display system. If participants only rely on two-dimensional information when they perceive gaze directions, no difference will be observed between the measurement results of the first and this experiment. The experiment should give us insight

about whether face-to-face type gaze communication can be realized using inexpensive monoscopic displays.

In the third experiment, stereoscopic images were used to transmit gaze because participants showed poorer performance under the monoscopic image condition than under the face-to-face condition. The experiment used stereoscopic images captured by two cameras whose distance was 6.5 cm, which is an average of participants' interpupillary distances. The distance between two cameras was not adjusted to each participant's interpupillary distance to assess the possibility of sharing the same stereoscopic images among users with different interpupillary distances.

The fourth and fifth experiments were conducted as preliminary studies to explore methods to design communication systems that enable remote communication that is close to face-to-face meetings. In the fourth experiment, we studied the influence of different interpupillary distances of users in designing communication systems. Stereoscopic images were captured by a stereo camera with lenses whose distance was 6.1 cm because measurement results for participants with interpupillary distance less than 6.3 cm had difficulty in judging the gaze direction. In this experiment, two participants who had an interpupillary distance smaller than 6.3 cm judged the gaze directions.

The fifth experiment examined an influence of head movement on the performance of gaze direction perception. The aim of this experiment was to assess the effects of the introduction of head tracking technologies to communication systems. Four participants joined the experiment.

Six participants, referred to as participant 1 . . . through participant 6, experienced the gaze conditions in a different sequence; and different experiments were conducted after more than two weeks to minimize the influence of adaptation.

## 2 Face-to-Face Experiment

The purpose of the face-to-face experiment is to measure the accuracy of participants' judgment of a real looker's gaze direction to find a quantitative base to compare different communication state.

## 2.1 Methods

**2.1.1 Participants.** Participants were two Japanese women, three Japanese men, and one French man, all in their twenties and familiar with the looker. All of them had normal or corrected-to-normal vision and could see stereo images correctly when their stereo vision was checked with a stereo test (Stereo tests, Stereo Optical co., Inc.). The same members participated in the first three experiments described in this paper.

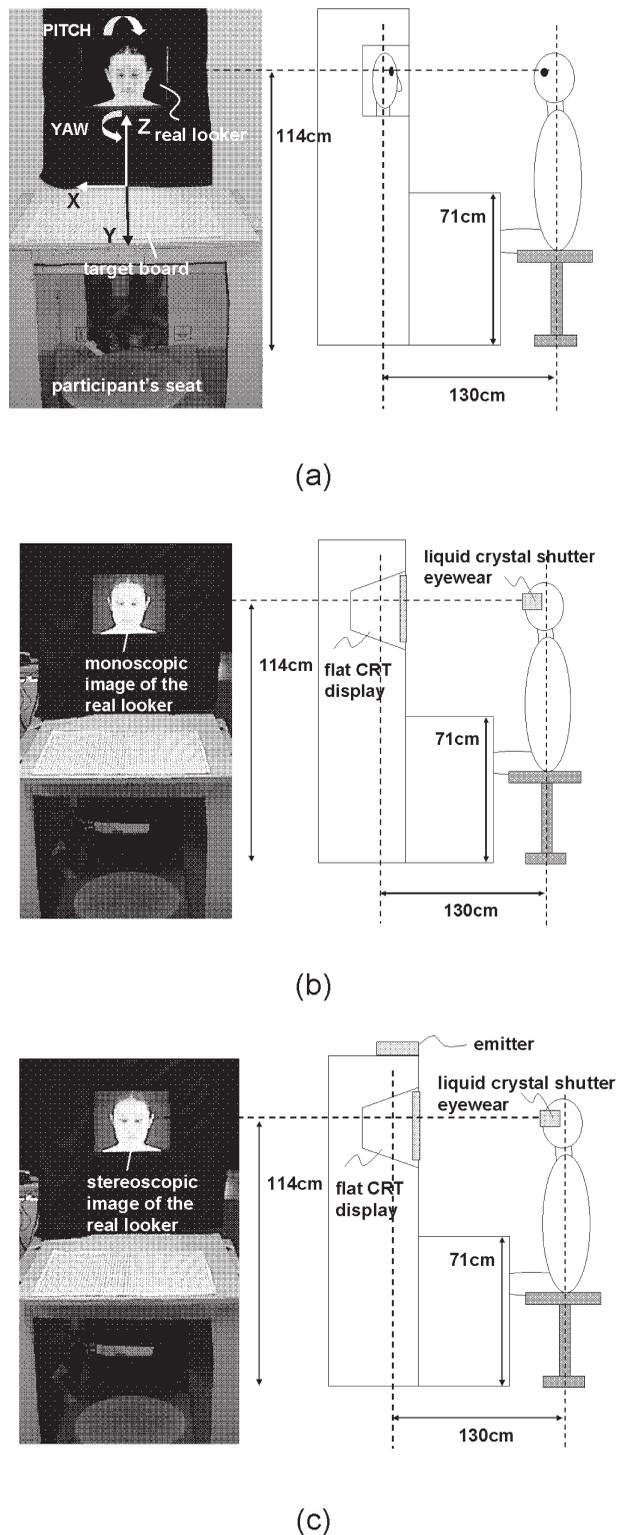
**2.1.2 Experimental Environment.** A participant and a looker were seated across a table and the looker fixated on targets on the table as shown in Figure 1a. The distance between the looker and the participants was 130 cm and the height of their eye positions was adjusted to 114 cm before starting the experiment. Their heads were not fixed during the experiment to recreate natural communication states because previous studies showed that when a looker's head was fixed and looked askance at targets, which is an unnatural condition, the judged locations were different from the results obtained under the condition in which gaze and target directions are parallel. For the purpose of investigation of natural communication state, head free conditions are appropriate. The height of the table was 71 cm. The looker was a Japanese woman who looked at the target board from a window of a box whose inside lighting was arranged to reduce shadows on the face and her head was not fixed. The size of the window was  $35.5 \times 23.2$  cm and the size of the looker's head was  $18.0 \times 23.0$  cm. The looker fixated through the window so that the face-to-face condition and display mediated conditions could be compared as accurately as possible. The targets were small black dots aligned in a grid at intervals of 1 cm. The grid had 54 dots in the  $x$  direction, which is from the left to the right hand side of the looker, and 40 dots in the  $y$  direction, which is in the depth direction of the looker, for a total of 2160 dots. The distance between the looker and the closest dot to her was 47 cm. The looker could distinguish all the dots with the naked eye. In this experiment, the looker fixated on 100 target points randomly.

The experiment was conducted in a room with ordinary lighting conditions and both the participants' and

the looker's heads were not fixed, in order to recreate conditions similar to those when two people discuss over a communication system. The participants closed their eyes while the looker prepared her fixation. The looker looked at one of the 2160 points in accordance with a prearranged random order. The participant then placed pins on the target board to mark points which they thought that she was looking at. The pins were numbered from 1 to 100 to show the relation between pins and fixations.

## 2.2 Results

**2.2.1 Comparison to Previous Research Results.** First, we compare our data to results obtained in a similar previous experiment to check whether our results are reasonable (Anstis et al., 1969). Since the measurement conditions are different, the data cannot be compared exactly but should show reasonable consistency. In the Anstis et al. experiment, the looker fixated seven numbered spots behind the participant's head, whereas the looker in our experiment looked down on a table. Participants answered judged locations with  $x$  and  $y$  on the target board but the results were decomposed into pitch and yaw angle components because Anstis et al. reported regression lines for the mean of six participants of yaw components. The coordinate system used in this analysis is shown in Figure 1a. The origin is at the looker's body center and height is 71 cm from the floor, positive  $x$  axis to the looker's right and positive  $z$  axis upward. A yaw component of the looker's gaze direction is defined as an angle between the positive  $y$  axis and a line drawn from a center of the looker to a fixated location. The yaw angle of the  $y$  axis is 0 and a positive yaw angle is in her left-hand side. A pitch component of her gaze direction is defined as an angle between the  $xy$  plane and a line drawn from her head location to a fixated target. The pitch increases from 0 to 90° when the looker shifts gaze direction from straight ahead to down on the floor. Figure 2 shows all the judged yaw component data and a regression line for the average data. Positive values of the horizontal axis in Figure 2 mean that the looker fixated her left side targets in Figure 1. A regression equation takes the form  $y = ax + b$ , where  $x$  is the yaw component of the looker's gaze direction and  $y$



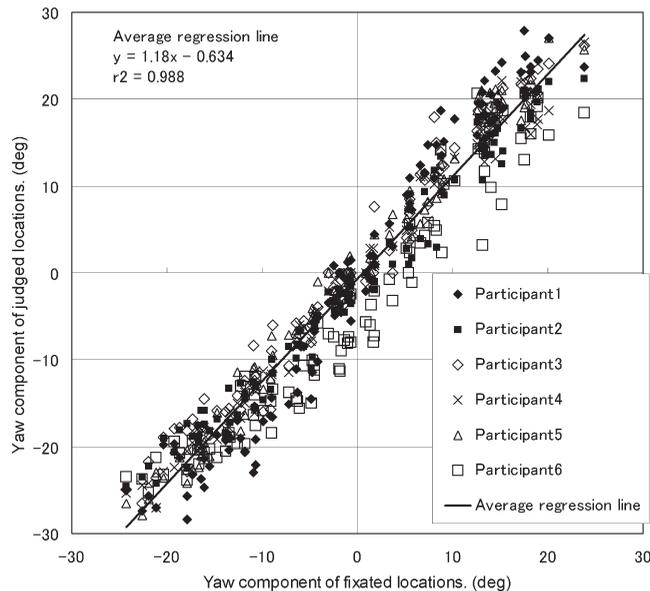
is the yaw component of the participant’s judgment. A regression line for the mean of six participants obtained in our experiment was  $y = 1.18x - 0.63$  as shown in Figure 2 and the result obtained by Anstis et al. was  $y = 1.50x - 0.05$ . Both regression coefficients are larger than 1.0, which implies that the gaze yaw angles are overestimated. Gibson and Pick (1963) Cline (1967) also reported the same tendency. The  $y$  intercept of our regression line has a minus sign as Anstis et al. reported. The only notable difference is the amount of the coefficients of the regression lines. Since Anstis et al. and our experiments were conducted under different conditions, we conclude that our experiment measured the accuracy of participants gaze direction perception properly.

**2.2.2 Gaze Perception Map.** Next, we explain the relation between positions that the looker fixated and the judgment of the participants in Figure 3. Here, an error vector is defined as a vector drawn from a fixated location to a judged location. The  $x$  and  $y$  axis correspond to the  $x$  and  $y$  axis shown in Figure 1 and the looker’s location is at  $(0, 0)$ . Figure 3 shows error vectors averaged over all the participants. Participants judged fixated locations much closer to the looker than the actual locations. The result revealed that we perceive a collaborator’s focus of attention much closer to the person than the correct targets.

### 3 Monoscopic Image Experiment

The purpose of this experiment is to measure the accuracy of participants’ judgment of gaze presented by a monoscopic image of the same looker as in the face-to-face experiment. If people judge collaborators’ gaze based on two-dimensional information, no difference should be observed between the face-to-face and the monoscopic image experiments. Then inexpensive

**Figure 1.** Experimental environments. (a) Face-to-face experiment. (b) Monoscopic image experiment. (c) Stereoscopic image experiment. The image was presented by a time-multiplexed stereoscopic CRT display with refresh rate 120 Hz.



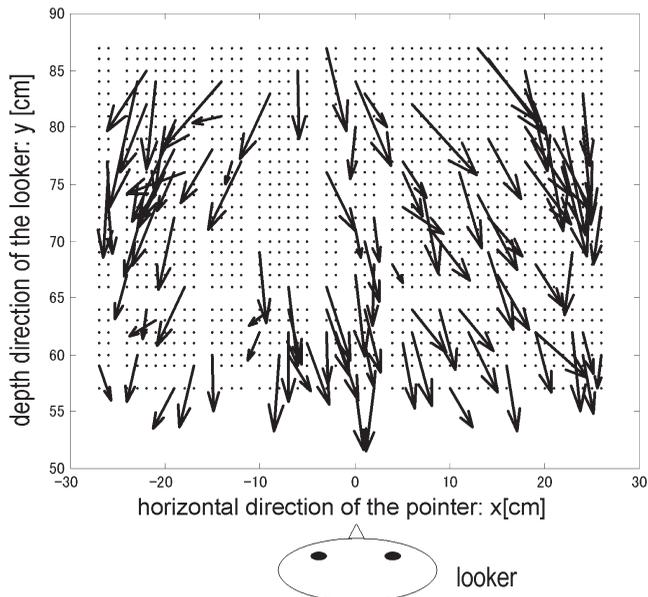
**Figure 2.** Yaw component data obtained under the face-to-face experiments.

monoscopic displays could be used to realize face-to-face-like remote gaze communication.

### 3.1 Experimental Environment

In this experiment, the real looker was replaced by a monoscopic image, as shown in Figure 1b. Still images of the looker were presented by a flat 19 inch CRT display. The display was placed behind the window that was used in the face-to-face experiment and part of the display area,  $30.7 \times 23.2$  cm, was shown from the window. The apparent size of the looker's head in the images was controlled so that it had the same apparent size as the control condition. The image was taken by a camera placed at the same position as the eye position of participants, which is 130 cm from the looker and at a height of 114 cm. The image resolution was  $1024 \times 768$  pixels and presented with 32 bit color. The size and the eye position of the image was arranged to be the same as the face-to-face condition. The experiment was conducted using the same box and room as those for the face-to-face experiment.

Participants were the same as in the face-to-face experiment but now monoscopic images were presented



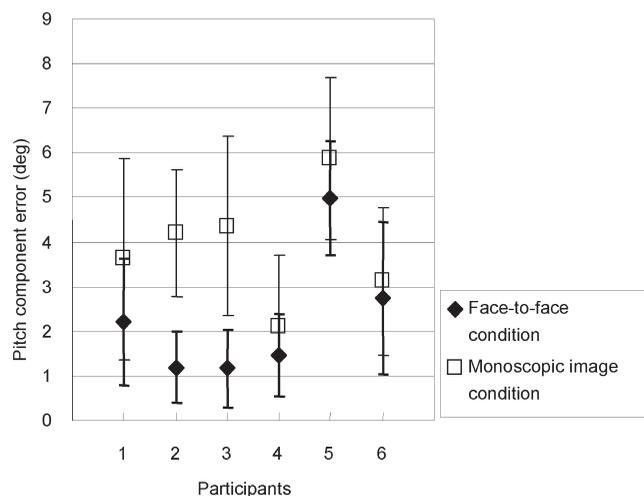
**Figure 3.** Error vectors obtained under the face-to-face condition.

through the window of the box. The distance between the participant and the looker's image was 130 cm and the height of the eye positions of participants was set to 114 cm before starting the experiments but their heads were not fixed. The participants wore stereo shutter glasses to make the comparison between the monoscopic and stereoscopic image conditions as accurate as possible.

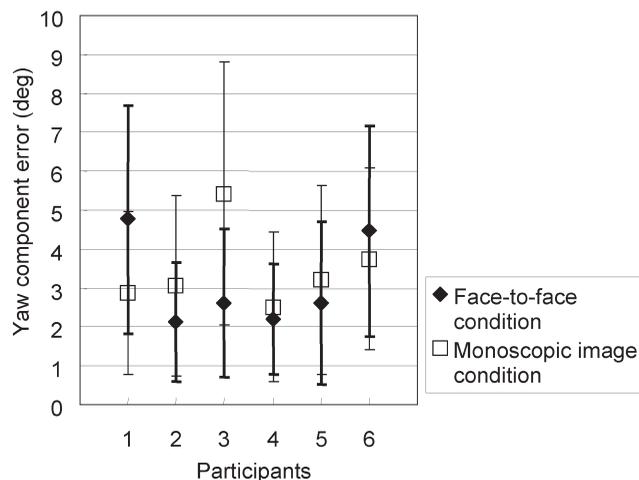
### 3.2 Results for Monoscopic Image Experiment

When pitch components of fixated and judged locations are compared, all the participants have larger errors under the monoscopic image condition than under the face-to-face condition and the difference is significant except participant 6 as shown in Figure 4 (two-tailed  $t$ -test:  $t_1(99) = 5.06$ ,  $p_1 < .001$ ,  $t_2(99) = 18.0$ ,  $p_2 < .001$ ,  $t_3(99) = 14.9$ ,  $p_3 < .001$ ,  $t_4(99) = 3.5$ ,  $p_4 < .001$ ,  $t_5(99) = 4.18$ ,  $p_5 < .001$ ,  $t_6(99) = 1.7$ ,  $p_6 > .05$ ). Here, a pitch component error is defined as an average of differences between pitch angles of fixated and judged locations.

However, consistent tendencies are not observed for



**Figure 4.** Difference between pitch components of actual regards and participants' judgment under the face-to-face and the monoscopic image conditions.



**Figure 5.** Difference between yaw components of actual regards and participants' judgment under face-to-face and monoscopic image conditions.

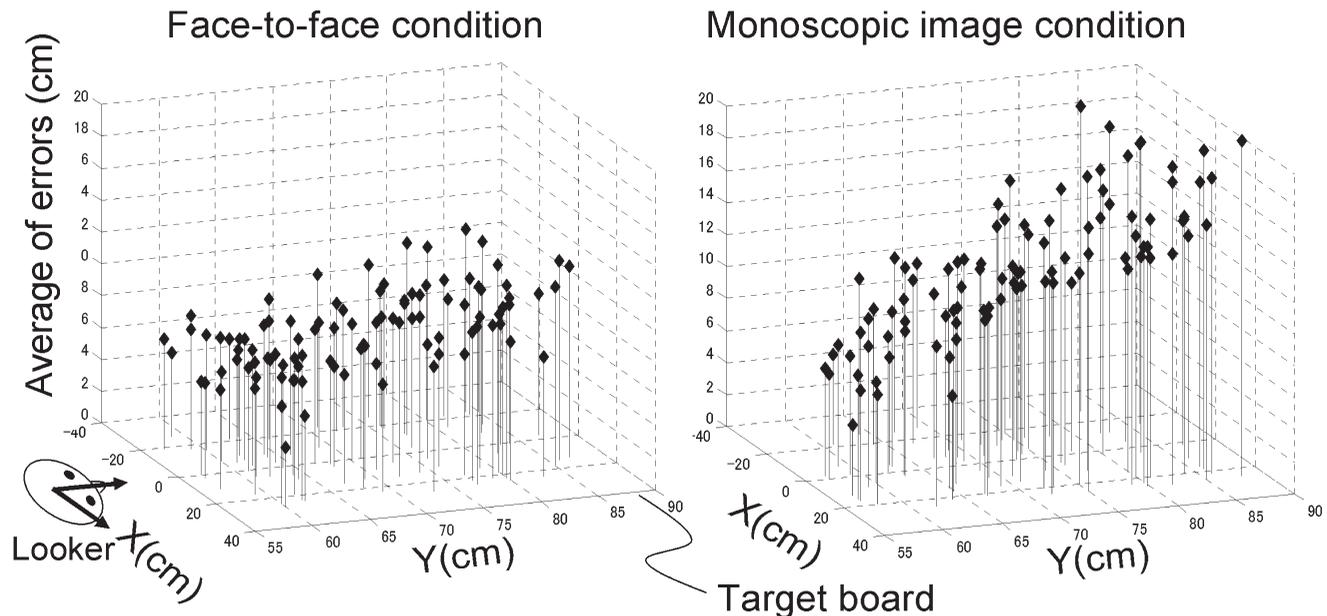
yaw component error in Figure 5. Participant 2 and participant 3 performed significantly better under the face-to-face condition than under the monoscopic image condition ( $t_2(99) = 8.88$ ,  $p_2 < .001$ ,  $t_3(99) = 2.33$ ,  $p_3 < .05$ ) but participant 1 and participant 6 performed better under the monoscopic image condition ( $t_1(99) = 9.19$ ,  $p_1 < .001$ ,  $t_6(99) = 4.83$ ,  $p_6 < .001$ ). From the analysis above we confirmed that monoscopic image made judgment of the pitch angle of the gaze direction difficult as we expected.

To show how this difference affects collaboration tasks intuitively, we plot errors for each stimulus in Figure 6. Here, the error is defined as a distance between a fixed and a judged location and the length of a stem in Figure 6 represents the size of average error taken over all the participants. The error is calculated in length instead of angle because it is useful to know how accurately we can share focus of attention on a working area in length. It can be seen that difference between the face-to-face and the monoscopic image condition is small when the looker looked at targets close to her but the difference becomes larger when fixated locations are at farther places. This suggests that the difference between the face-to-face and the monoscopic image conditions becomes acute when the work areas become

large. Also, the perception of the gaze direction pitch component depends on the display technology but the yaw component is independent of display technology. Therefore, we will analyze pitch component errors to compare results obtained under different experimental conditions. Also, participant 6 may have a personal tendency of not using or difficulty in obtaining 3D information when she judges the looker's gaze direction because she did not show significant difference between the face-to-face and the monoscopic image conditions. This personal difference in 3D information acquisition will be examined in our experiments below. These results show that participants cannot judge the looker's gaze correctly with 2D information, and therefore 3D information is required to achieve realistic gaze communication.

#### 4 Stereoscopic Image Experiment

The monoscopic image experiment revealed that the two-dimensional information is not sufficient to achieve face-to-face like gaze communication. In this experiment, we measure the accuracy of participants' judgments of gazes presented by a stereoscopic image of the same looker as in the face-to-face experiment.



**Figure 6.** Errors for each target point. The length of the bar shows an average of absolute difference between fixated and judged locations.

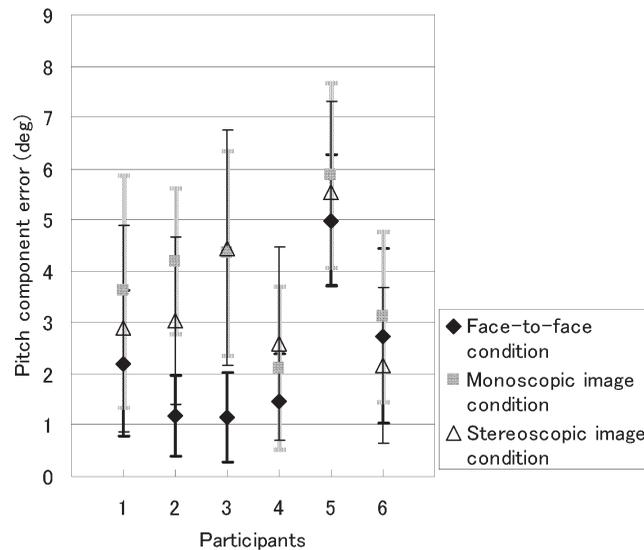
#### 4.1 Experimental Environment

The experimental apparatus and the procedure were the same as in the monoscopic image experiment as shown in Figure 1c. The stereoscopic image resolution was lower than that of the monoscopic image,  $640 \times 368$  pixels, because the two images, one for the left eye and the other for the right eye, were to be displayed on the same display and they were stretched to cover the display area in the stereo display mode. The image was presented by a time-multiplexed stereoscopic CRT display with refresh rate 120 Hz (Crystal EYES PC, StereoGraphics Corporation). The looker's size seen through the stereo shutter glasses was adjusted to be the same size as for the real looker. Before capturing the image, interpupillary distances of all the participants were measured with a Towa PD METER (Towa PD METER PD-82, Towa Medical Instruments Co., LTD). Participant 1 wore glasses in the measurement because he was very nearsighted and could not see the target point in the PD METER without them. The interpupillary distances of participants were 6.8, 6.7, 6.1, 6.4, 6.6, 6.2 cm and the average was 6.5 cm. Two cameras were placed 6.5 cm apart when stereo images were shot and the stereo camera distance was not adjusted to

participants' interpupillary distances to evaluate influence of using the same stereoscopic images for different users. Here, we define a stereo camera distance as a distance between centers of two lenses of cameras that are used in shooting stereoscopic images.

#### 4.2 Results for Stereoscopic Image Experiment

On average, participants perceived gaze directions most accurately under the face-to-face condition and perceived result was worst under the monoscopic image condition as expected. The average pitch angle error for the face-to-face condition is  $2.29^\circ$  and the error for the stereoscopic and monoscopic image conditions is  $3.44^\circ$  and  $3.88^\circ$ . There are significant differences between the face-to-face and the stereoscopic image condition ( $t(99) = 11.4, p < .01$ ), and between stereoscopic and monoscopic image conditions ( $t(99) = 3.95, p < .01$ ). However, data for each participant did not follow the expected order. Participant 1 ( $t(99) = 2.3, p < .05$ ), participant 2 ( $t(99) = 5.6, p < .01$ ), and participant 6 ( $t(99) = 4.2, p < .01$ ) perceived gaze directions more precisely under the stereoscopic image condition than



**Figure 7.** Pitch component errors obtained under the face-to-face, the stereoscopic image, and the monoscopic image conditions.

under the monoscopic image condition as shown in Figure 7. However, participant 3 ( $t(99) = 0.29, p > .05$ ), participant 4 ( $t(99) = 2.0, p = .05$ ), and participant 5 ( $t(99) = 1.37, p > .05$ ) did not show significant difference. The image resolution difference may have some influence on the results but there is a research result that shows little influence of the resolution of images on perception of gaze directions (Sato & Matsuzaki, 2000). Participant 6 perceived the gaze under the stereoscopic condition better than under the face-to-face condition. The result is anomalous because under the face-to-face condition, people judge gaze directions with a systematic shift toward the looker but participant 6 did not show the shift under the stereoscopic image condition. The result suggests that the stereoscopic image did not provide correct stereo information to participant 6. Since participant 6 had a narrower interpupillary distance, 6.2 cm, compared to other participants, she might have observed distorted images. Participant 3 who has the narrowest interpupillary distance of 6.1 cm did not show significant difference between monoscopic and stereoscopic image conditions; and his difference between pitch angle errors obtained under stereoscopic image and the face-to-face condition was the largest of all the participants. These results suggest that interpupil-

lary distances had influence on perception of gaze directions in this experiment. We will investigate this issue later in this paper.

Table 1 shows regression lines,  $y = ax + b$ , for pitch components of perceived gaze directions. Here,  $x$  is the pitch component of the looker's gaze direction and  $y$  is the pitch component of the participant's judgment. The regression coefficients and the  $y$  intercept are determined by fitting the equation to data with the least square method. The regression coefficients under the face-to-face condition are larger than or equal to 1 except participant 5 who perceived gaze directions differently than others. The participant 5 has the coefficients less than 1 and  $y$  intercept larger than 10 under all the conditions. The same tendency can be found in Table 1 for participant 3 under the monoscopic image and the stereoscopic image conditions. Since the difference between the interpupillary distance of participant 3, 6.1 cm, and the stereo camera distance, 6.5 cm, is the largest of the six participants, the small coefficient and the large  $y$  intercept tend to show up when participants have difficulty in obtaining 3D information in judging gaze direction. Therefore, it is suspected that participant 5 mainly used 2D cues in judging gaze direction even if 3D information was given. Under the face-to-face condition, five participants have coefficients larger than 1 and under the stereoscopic image condition, the number reduced to three. Under the monoscopic image condition, only participant 2 has a coefficient larger than 1. The coefficient reflects how far the displayed image is close to the face-to-face condition. Also, under the stereoscopic image condition, the participants who have larger interpupillary distance than 6.4 cm have the coefficients larger than 1. This tendency suggests that participants with larger interpupillary distance judged the stereoscopic image condition closer to the face-to-face conditions than the participants with narrower interpupillary distance.

We summarize the trends of the coefficients in Table 2. Data for all the participants were used to calculate a regression line. The coefficients for yaw component do not show clear trends but the coefficients for pitch component becomes close to 1 when the given information becomes closer to the face-to-face condition. Also, the  $y$

**Table 1.** *Interpupillary Distance and Regression Lines for Pitch Components\**

PN	IPD (cm)	Face to face	Stereoscopic	Monoscopic
1	6.8	$y = 1.2x - 3.8$	$y = 1.1x + 0.96$	$y = 0.99x + 3.8$
2	6.7	$y = 1.1x - 3.6$	$y = 1.3x - 4.8$	$y = 1.1x + 2.0$
3	6.1	$y = 1.1x - 2.7$	$y = 0.62x + 15$	$y = 0.72x + 12$
4	6.4	$y = 1.0x + 0.44$	$y = 1.1x - 1.6$	$y = 0.91x + 4.5$
5	6.6	$y = 0.67x + 14$	$y = 0.66x + 15$	$y = 0.58x + 18$
6	6.2	$y = 1.1x - 0.98$	$y = 0.94x + 3.7$	$y = 0.89x + 6.1$

\*The distance between two cameras was 6.5 cm. PN: participant number, IPD: interpupillary distance.

**Table 2.** *Regression Lines for Pitch and Yaw Components*

Conditions	Regression lines for pitch component	Regression lines for yaw component
Face-to-face	$y = 1.04x + 0.608$	$y = 1.18x - 0.634$
Stereoscopic image	$y = 0.949x + 4.75$	$y = 1.16x - 0.639$
Monoscopic image	$y = 0.861x + 7.70$	$y = 1.21x + 1.13$

intercepts for the pitch component get closer to 0 when the condition becomes closer to the face-to-face condition.

## 5 Stereoscopic Image Experiment for Participants with Narrow Interpupillary Distances

The stereoscopic image experiment suggested that the perception of gaze direction is affected by interpupillary distances of the participants. In this experiment, we will measure the influence of the stereo camera distance on the perception of the gaze direction presented by stereoscopic images.

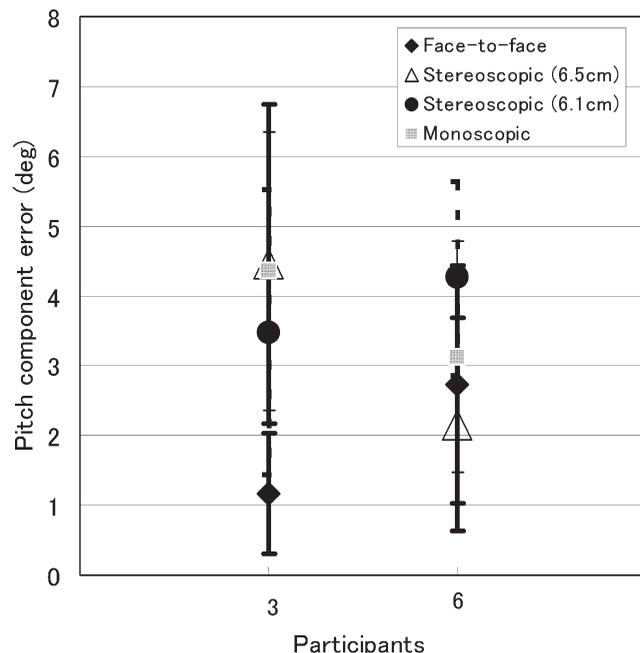
### 5.1 Experimental Environment

In the previous experiment, the participants with larger interpupillary distances than the stereo camera distance did not show any anomalous results. Therefore, in this exploratory experiment, stereoscopic images were captured using two cameras separated by the interpupil-

lary distance of participant 3, which is 6.1 cm, because he had the narrowest interpupillary distance of all the participants. The experiment will provide insight into possibility of using stereo images that are shot with a stereo camera with a distance that is fixed to be the narrowest interpupillary distance of the expected user of the system.

### 5.2 Results

As expected, the anomalous result that was obtained with participant 6 was corrected by making the stereo camera distance to 6.1 cm as shown in Figure 8 and the improvement is significant ( $t(99) = 5.47, p < .01$ ). When the stereo camera distance was larger than the interpupillary distance of participant 6, the error was smaller under the stereoscopic image condition than under the face-to-face condition. However, when the camera distance was modified to 6.1 cm the error under the stereoscopic image condition became larger. Also, the error for participant 3 became smaller when the camera distance was matched to his interpupillary dis-



**Figure 8.** Pitch component errors under different stereo camera distance conditions.

tance. Under the camera distance 6.5 cm condition, the difference of pitch component errors are insignificant between monoscopic and stereoscopic image conditions ( $t(99) = 0.29, p > .05$ ). However, the difference became significant when the camera distance was tuned to 6.1 cm ( $t(99) = 2.89, p < .01$ ).

The regression coefficients and  $y$  intercepts for both participants also became closer to the values obtained under the face-to-face condition as shown in Table 3.

## 6 Face-to-Face Experiment with Fixed Head

The purpose of this experiment is to identify the influence of active head movements of the participants in gaze direction perception. This experiment should give insight into how far the introduction of head tracking technologies can improve gaze direction perception.

## 6.1 Experimental Environment

The apparatus and the looker were the same as in the face-to-face condition. The participants' heads were fixed by a chin rest but the looker's head was not fixed. Participants 2, 3, 4, and 5 joined this exploratory experiment.

## 6.2 Results

The restriction of head movement increased the pitch angle error for participants 2 and 4 but the error became smaller for participants 3 and 5 as shown in Figure 9. The pitch angle error for participant 2 under the head-free condition was  $1.18^\circ$  and it became  $1.71^\circ$  when the head was fixed ( $t(99) = 2.3, p < .05$ ). The error also increased for participant 4 from  $1.46^\circ$  to  $2.08^\circ$  ( $t(99) = 3.5, p < .01$ ). However, the error for participant 3 decreased from  $1.16^\circ$  to  $0.65^\circ$  ( $t(99) = 3.2, p < .01$ ) and from  $4.99^\circ$  to  $3.72^\circ$  for participant 5 ( $t(99) = 11, p < .01$ ). The result suggests large personal differences in using motion parallax information to judge gaze direction. When data obtained under the head-fixed and stereoscopic image conditions are compared, all the participants performed better under the head-fixed condition as shown in Figure 9 ( $t_2(99) = 5.4, p_2 < .01, t_3(99) = 12, p_3 < .01, t_4(99) = 2.17, p_4 < .05, t_5(99) = 11, p_5 < .01$ ). Here, stereoscopic image data for participant 3 was captured by a stereo camera with an interpupillary distance of 6.1 cm.

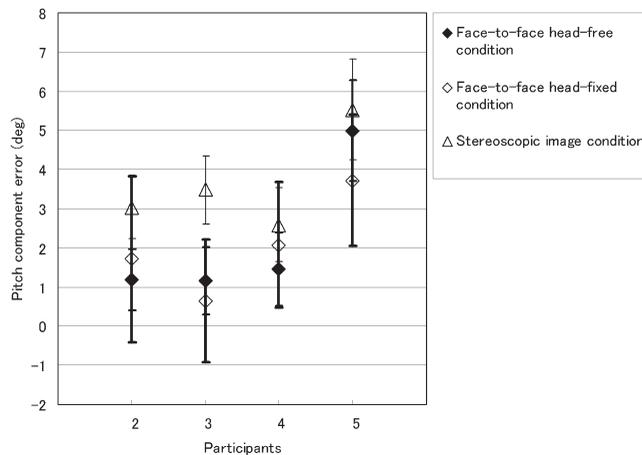
The standard deviations ( $SD$ ) of the pitch component errors for the head-free condition are always smaller than the  $SD$  obtained under the head-fixed and stereoscopic image conditions as shown in Table 4. However, clear trends are not observed for  $SD$ s obtained under the head-fixed and the stereoscopic image conditions. The results suggest that introduction of head tracking systems to communication systems will not improve the quality of remote communication drastically and its effect depends on users.

## 7 Discussion

The average data show that the accuracy of gaze direction perception of participants is the highest under

**Table 3.** Influence of the Distance Between Stereo Cameras on Gaze Direction Perception

PN	IPD (cm)	Face to face	Stereoscopic (Camera: 6.1 cm)	Stereoscopic (Camera: 6.5 cm)	Monoscopic
3	6.1	$y = 1.1x - 2.7$	$y = 0.93x + 5.5$	$y = 0.62x + 15$	$y = 0.72x + 12$
6	6.2	$y = 1.1x - 0.98$	$y = 1.1x + 1.9$	$y = 0.94x + 3.7$	$y = 0.89x + 6.1$

**Figure 9.** Influence of head fixation on the gaze direction perception.

the face-to-face condition, followed by the stereoscopic image condition, and the lowest under the monoscopic image condition as expected. The average pitch angle error for the face-to-face condition is  $2.29^\circ$ , and the error for the stereoscopic and monoscopic image conditions are  $3.44^\circ$  and  $3.88^\circ$ . There are significant differences between the face-to-face and the stereoscopic image condition ( $t(99) = 11.4, p < .01$ ), and between stereoscopic and monoscopic image conditions ( $t(99) = 3.95, p < .01$ ). This result suggests that a certain improvement of gaze perception can be expected by using stereoscopic images rather than monoscopic images, even if the same stereoscopic images are shared between people. Further improvement was observed under the stereoscopic image condition when the images were captured with two cameras whose distance was smaller than or equal to the participants' interpupillary distances. However, only participants 2 and 3 followed the expected order when individual's data is examined. The participant 6 had smaller errors in per-

ceiving the gaze directions under the stereoscopic image condition than under the face-to-face condition when the camera distance was 6.5 cm. However, the unnatural result was corrected by adjusting the stereo camera distance to his interpupillary distance. Since participants with interpupillary distances larger than the stereo camera distance had the coefficients larger than 1, which is a trend observed under the face-to-face condition, the stereo camera distance should be adjusted to be smaller than the expected users of the stereoscopic images.

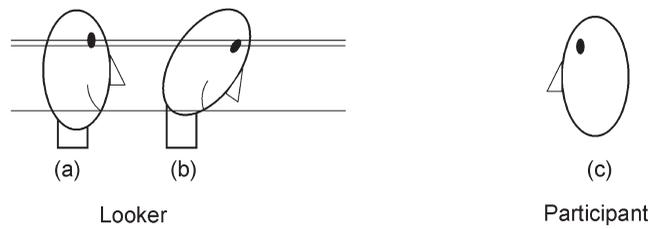
Also, the pitch component errors were gaze transmission media sensitive, but the yaw component errors were robust when the gaze perception capability under different conditions is compared using average angle errors as shown in Figures 4 and 5 and regression lines as shown Table 2.

The regression coefficients in Table 1 reflect difference in transmission methods of gaze directions. On average, a coefficient obtained under the face-to-face condition is larger than 1 and it is smaller than 1 under the monoscopic image condition. Participants with interpupillary distance larger than or equal to 6.4 cm have coefficients larger than 1.0. However, the coefficients for participant 5 under the face-to-face and stereoscopic image conditions are smaller than 1 even though he has an interpupillary distance larger than 6.4 cm. Also, the  $y$  intercept for the participant is the largest of all the participants. This suggests that he used cues that are independent from presentation methods. Therefore, it is suspected that he mainly used 2D cues even under the face-to-face condition. When participants judge orientation of the looker's head, it is possible to use many different cues. For example, a distance between eyes and mouth can be used as one of the 2D cues because when the looker tilts her head as shown in Figure 10a, b the distance becomes smaller when the participant (Figure

**Table 4.** Interpupillary Distance and Regression Lines for Pitch Components\*

PN	SD for head-free condition	SD for head-fixed condition	SD for stereoscopic image condition (6.5 cm)
2	0.79	2.1	1.6
3	0.87	1.6	2.1
4	0.94	1.6	1.9
5	1.3	1.7	1.8

\*The distance between two cameras was 6.5 cm. PN: participant number, IPD: interpupillary distance.



**Figure 10.** Distance between an eye and a mouth looks different when the looker tilts her head.

10c) observes the looker. If this kind of 2D cue is mainly used, there should be small difference between data obtained under different experimental conditions. Also, it is difficult to estimate the location of the looker based on the 2D cues. Participant 5 may have mainly used this kind of 2D information because he showed small differences under different experimental conditions and has the largest  $y$  intercept of all the participants.

When we treat participant 5's data as a special case, participants with interpupillary distances larger than or equal to 6.4 cm have coefficients larger than 1.0 and participants with interpupillary distances less than 6.4 cm have coefficients smaller than 1.0. The relation between the coefficients and the interpupillary distances can be explained by examining actual and judged change in the gaze pitch angles because the coefficient is defined as  $a = \Delta y / \Delta x$ , where  $\Delta y$  is the shift in judged pitch angle and  $\Delta x$  is the pitch angle change of actual gaze direction.

Firstly, we calculate a distance between nose locations shot by the left and the right camera as shown in Figure 11. When a tip of the nose  $p_{\text{nose}}$  is shot by the left cam-

era, it is displayed at  $n_l$  and the nose shot by the right camera is displayed at  $n_r$ . From the relation

$$D_{\text{cameras}} \cdot D_{\text{imageNose}} = (D_{\text{noseDisp}} + D_{\text{dispCamera}}) \cdot D_{\text{noseDisp}}$$

$D_{\text{imageNose}}$  can be calculated as

$$D_{\text{imageNose}} = \frac{D_{\text{noseDisp}}}{D_{\text{noseDisp}} + D_{\text{dispCamera}}} \times D_{\text{cameras}}$$

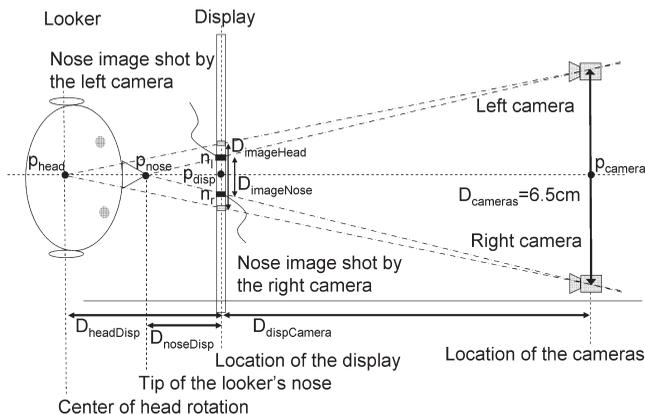
Here,  $D_{\text{cameras}}$  is the distance between the left and right cameras, and  $D_{\text{imageNose}}$  is the distance between  $n_l$  and  $n_r$  on the display.  $D_{\text{dispCamera}}$  is the distance between the display and the camera location  $p_{\text{camera}}$  and  $D_{\text{noseDisp}}$  is the distance between the tip of the nose and the display at  $p_{\text{disp}}$ . Under our experimental condition,  $D_{\text{imageNose}}$  is 0.441 cm with values  $D_{\text{cameras}} = 6.5$  cm,  $D_{\text{noseDisp}} = 8$  cm,  $D_{\text{dispCamera}} = 110$  cm. In the same way, distance between head locations shot by the left and the right cameras,  $D_{\text{imageHead}}$  can be calculated as follows

$$D_{\text{imageHead}} = \frac{D_{\text{headDisp}}}{D_{\text{headDisp}} + D_{\text{dispCamera}}} \times D_{\text{cameras}}$$

$D_{\text{imageHead}}$  is 1.00 cm with  $D_{\text{headDisp}} = 20$  cm. Here,  $D_{\text{headDisp}}$  is a distance between the head and the display locations.

Next, we examine a case when a participant with an interpupillary distance narrower than the camera distance judges the presented images. If the interpupillary distance,  $D_{\text{interp}}$ , is 6.1 cm, the judged head position  $p'_{\text{head}}$  is farther than the actual head location  $p_{\text{head}}$ . Using the relation

$$D_{\text{interp}} \cdot D_{\text{imageHead}} = (D'_{\text{headDisp}} + D'_{\text{dispCamera}}) \cdot D'_{\text{headDisp}}$$



**Figure 11.** Shooting environment for the stereoscopic images with two cameras with a stereo camera distance with 6.5 cm.

judged distance between the head and the display locations,  $D'_{headDisp}$ , is calculated to be 22 cm. This means that the looker's location is perceived as 2 cm farther than the actual location. The location of the nose is also perceived to be 1 cm farther than the actual location and the distance between the judged nose location and  $p_{disp}$ ,  $D'_{noseDisp}$  is 119 cm.

Now assume that the real looker rotates her head by  $\Delta\theta$  in a vertical direction at a location  $p_{head}$  and the change in the nose location is  $\Delta z$ . The relation can be approximated as

$$\Delta z = (p_{head} - p_{nose}) \times \Delta\theta$$

and the actual nose location shift  $\Delta z$  is displayed as  $\Delta z_{display}$  on the display. When a participant with a narrow interpupillary distance watches the change  $\Delta z_{display}$  on the display, he/she thinks that the looker rotated her head at a farther place,  $p_{head}$ , than the actual location. Therefore, the head rotation angle of the perceived looker  $\Delta\theta'$  is estimated to be smaller than the real looker's angle  $\Delta\theta$  and the ratio is calculated as follows

$$\frac{\Delta\theta'}{\Delta\theta} = \frac{D_{dispCamera} \times D_{interP} + D_{headDisp}(D_{interP} - D_{cameras})}{D_{dispCamera} \times D_{cameras}}$$

Since the coefficient of the regression line shows the ratio between actual and perceived head rotation angles, the equation above should approximate the coefficients obtained in our experiments. According to the equa-

tion, the ratio is 1 when the camera distance  $D_{cameras}$  is adjusted to a participant's interpupillary distance  $D_{interP}$ . The ratio is larger than 1 if an interpupillary distance is larger than the camera distance. These relations are the same as our experimental data shown in Table 1 but there is a difference between the theoretical boundary  $D_{cameras} = 6.5$  cm and the measured boundary 6.4 cm. The calculated ratios for interpupillary distances 6.8 cm and 6.7 cm are 1.1 and 1.0 and the measured coefficients are 1.1 and 1.3. The ratio for interpupillary distances 6.1 cm and 6.2 cm are 0.93 and 0.95, and the coefficients obtained by our experiments are 0.62 and 0.94 as shown in Table 1. The equation can recreate the trend observed in our experiment but cannot calculate the coefficients exactly. Fixation of the looker's head may lead to experimental results that are closer to the calculated ratio. However, special cautions are required in designing such experiments because it is known that when a looker watches targets askance because of the head fixation the results are different from experiments conducted under head-free conditions.

The fixation of the participants head had a large personal difference and consistent deterioration of the gaze direction perception was not observed. Since the cues used to perceive gaze direction may differ from individual to individual, it is possible that not all participants used motion parallax information in judging gaze direction.

The difference between the looker's fixated points and the participants' judgments increase when the points are far from the looker in the depth direction as shown in Figure 6. The influence of the different presentation methods on the errors is small when the targets are closer to the looker. However, the errors increase rapidly when the location becomes far from the looker under the monoscopic image condition.

Our experimental results provided us two insights into designs of communication systems. They are personalization of images and the anisotropic nature of different presentation methods. A stereo camera's lens distance should be adjusted to each user of the stereoscopic image. If the personalization is not possible and users of the images are known, the stereo camera distance should be adjusted to the narrowest interpupillary distance of expected users. Also, if the collaboration

tasks are done at locations that are close to the looker, a monoscopic image can be used. If the working area is extended into the depth direction, it is desirable to use a stereoscopic image.

## 8 Conclusion

We conducted experiments to answer the following questions. How is visual information exchanged in face-to-face communication modified when a communication system is introduced? How can the difference be measured quantitatively? Can the results of such a measurement be used to design and compare new communication systems? How can the system be improved?

The participants understood gaze directions most accurately under the face-to-face condition and least accurately under the monoscopic image condition. The performance was improved when stereoscopic images were captured with two cameras separated by an observer's interpupillary distance. Display conditions had little influence on the perception of the gaze directions when the looker fixed targets closer to her. However, the influence became large when the fixed locations were far from the looker into the depth direction. Also, the perception of the pitch component of gaze directions depends on the presentation methods but the yaw component does not depend on the presentation methods.

The gaze communication capabilities of different displays can be compared using a regression lines for the pitch component of the gaze directions. Also, the similarity between the technology mediated gaze communication and the face-to-face communication can be evaluated by comparing regression coefficients for the pitch component of gaze perception measurement results. If presented gazes are similar to gazes under face-to-face conditions, the average regression coefficients should be greater than or equal to 1.0.

The results can be used to design new communication systems. If a user only needs to communicate horizontal gaze information and the objects to be shared are placed at locations that are close to the looker, inexpensive monoscopic display can be used to develop the system. However, if a user has to convey gazes into depth directions, expensive stereoscopic display is desirable. If

several users want to share accurate gaze information, technologies to capture and display personalized stereoscopic images are required.

We developed a measurement method to determine whether a display has gaze communication capability that is similar to the face-to-face condition. The measurement method can be used for displays with different shapes and technology.

## Acknowledgments

We thank members of Tachi Laboratory for valuable discussions and suggestions. We also thank participants who volunteered to support experiments. This research was supported by the Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Corporation (JST).

## References

- Anstis, S. M., Mayhew, J. W., & Morley, T. (1969). The perception of where a face or television 'portrait' is looking. *American Journal of Psychology*, *82*, 474–489.
- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, UK: Cambridge University Press.
- Cline, M. G. (1967). The perception of where a person is looking. *American Journal of Psychology*, *80*, 41–50.
- Fish, R. S., Kraut, R. E., & Chalfonte, B. L. (1990). The Video Window system in informal communications. *Proceedings of CSCW 90*, 1–11.
- Fish, R. S., Kraut, R. E., & Root, R. W. (1992). Evaluating video as a technology for informal communication. *Proceedings of the CHI 92*, 37–48.
- Garau, M., Slater, M., Bee, S., & Sasse, M. A. (2001). The impact of eye gaze on communication using humanoid avatars. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 309–316.
- Gibson, J. J., & Pick, A. D. (1963). Perception of another person's looking behavior. *American Journal of Psychology*, *76*, 386–394.
- Ishii, H., Kobayashi, M., & Jonathan, G. J. (1993). Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems*, *11*(4), 349–375.
- Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387* (6635), 767–768.

- Lee, S. P., Badler, J. B., & Badler, N. I. (2002). Eyes alive. *Proceedings of ACM SIGGRAPH 2002*, 637–644.
- Sato, T., & Matsuzaki, N. (2000). Effects of viewing distance and image resolution on gaze perception. *Proceedings of Workshop on Virtual Reality*, 127–131 (in Japanese).
- Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253, 265–266.
- Tachi, S., & Arai, H. (1997). Design and evaluation of a visual display with a sensation of presence in tele-existence system. *Journal of Robotics and Mechatronics*, 220–230.
- Taylor, M. J., & Rowe, S. M. (2000). Gaze communication using semantically consistent spaces. *Proceedings of the CHI 2000*, 400–407.
- Vertegaal, R. (1999). The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. *Proceeding of the CHI 99*, 294–301.

Copyright of *Presence: Teleoperators & Virtual Environments* is the property of MIT Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.